View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 57 (2014) 930 – 939

Energy Procedia

2013 ISES Solar World Congress

Solar energy research in Ibero-America, a citation mining approach

Karla G. Cedano[a]*, Karla F. Ricalde[b], J. Antonio del Río[c], Manuel Martínez[c]

[a]*Centro Lavín para el Desarrollo de Innovación y Transferencia Tecnológica, Av. Domingo Diez 1003 3er piso, Cuernavaca, 62250, México*

[b]*Facultad de Ciencias, UNAM, Av. Universidad 3000, México D.F., 04150, México*

[c]*Instituto de Energías Renovables, UNAM, Priv. Xochicalco s/n, Temixco, 62250, México*

Abstract

In this paper we present an analysis of all the research papers published on journals registered in the Web of Science, with the phrase "solar energy" in its title, abstract or keywords that have at least one author with address in Ibero-America. We present the results of citation mining applied to all such records published between 2002 and 2012. This analysis characterizes the behavior of the scientific production on solar energy in the most prolific countries in this region.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Selection and/or peer-review under responsibility of ISES.

Keywords: Solar energy; prospective analysis; scientometrics

1. Introduction

The research on solar energy in Ibero-America has increased both in quantity and quality of publications. However, to define an optimal development strategy it is important to make an accurate assessment. This is done by taking into account either the authors or the institutions involved in the publication.

In this work we analyze all the paper registered in the Web of Science (Thomson Reuters) that appear under the search of the phrase "solar energy". The search was done taking into consideration titles, abstracts and key words of papers published worldwide between 2002 and 2012. Then, we selected Ibero-

* Corresponding author. Tel.: +52-777-230-5562.
E-mail address: karla.cedano@inno-ba.mx

1876-6102 © 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Selection and/or peer-review under responsibility of ISES.
doi:10.1016/j.egypro.2014.10.075

American countries with an individual production larger than or equal to 1% of the worldwide total. The search was done in June 2013.

It is noteworthy that WoS is the one of the most important data bases of scientific information in the world and that there is a bias towards papers written in English. Notwithstanding, we consider the sample to be significant for the study of the impact and pertinence of this research topic.

Recently, the analysis of citation mining [1], has been applied to study the characteristics of Mexican science in two of the most important journals Nature and Science [2] and to depict Ibero-American science (Russell et al. 2007).

The work is organized as follows: first, we present the method used for the analysis of the citation mining. In section 3 we applied a logistic method in order to foresee the evolution tendency of the area in the selected countries. In section 4 we show the bibliometric results, and in section 5 the results of the citation mining in the mentioned articles are provided. Finally we present our comments and interpretations of the results

2. Methodology

The citation mining methodology is based on the application of a combination of bibliometric techniques and text mining for the analysis of the bibliographic data [1] [3]. In this case study, the objective has been defined as the research papers on "solar energy" written between 2002 and 2012 excluding proceedings, reviews and editorials. The search resulted in 5,866 papers. Table 1 shows the countries with the most significant contributions to the area, those with a contribution of at least 1%.

From this table we selected the countries in Ibero-America: Spain, Brazil and Mexico, with overall contributions of 4.9, 1.77 and 1.06 per cent, respectively; adding up to 7.74%. We downloaded three sets of records, one per country.

The software tool that our research group has developed for this purpose analyzed each set [3]. Whilst the bibliometric stage is exclusively done by counting similar data from different fields on such bibliographic records, the text mining stage uses an entropy based algorithm to find the most relevant words in the abstracts of the records. This algorithm is based on the research done by Ortuno et al [4]. The distance between two occurrences of a particular word occurring in the text of an abstract was compared to the standard deviation of all words in all abstracts. A normalized standard deviation higher than 1 indicates that the distribution of the word within a particular abstract is not random allowing us to determine which words or strings of words can be considered relevant for that particular text. The reasoning behind this assumption is that the standard deviation is an analogous indicator to entropy [5] and can sometimes play a role as a measure of order (or disorder). The advantage of this particular technique is that it does not require a labor-intensive revision of individual words to extract the keywords from a text but rather provides a ready-made list of the most frequently occurring words and strings of words whose distribution within a text is not random and, therefore, likely to be significant. This technique has been recently used to analyze topics on highly visible science [6].

We based our prospective analysis on the notion that all biological, social and economic systems within a closed space have a natural cycle of birth, growth and saturation [7]. Hence if a time series has shown in the past a "natural growth", then its cumulative growth in time must have the shape of an "S" curve, also known as the logistic function. We applied a logistic regression, which is a canonical link function, meaning that parameter estimates under logistic regression are fully efficient, and tests on those parameters are better behaved for small samples. So we analyzed the scientific production over time of the different sets we classified in the first place, and applied our interactive logistic fit algorithm to it.

Table 1. Papers published by country (individual contribution greater than 1% of total of paper)

Countries	Papers	%
USA	1172	19.98%
PEOPLES R CHINA	831	14.17%
INDIA	399	6.80%
JAPAN	363	6.19%
GERMANY	323	5.51%
SPAIN	288	4.91%
TURKEY	275	4.69%
FRANCE	239	4.07%
ITALY	215	3.67%
ENGLAND	195	3.32%
CANADA	182	3.10%
SWITZERLAND	181	3.09%
TAIWAN	168	2.86%
SOUTH KOREA	165	2.81%
AUSTRALIA	146	2.49%
GREECE	114	1.94%
BRAZIL	104	1.77%
SWEDEN	103	1.76%
ISRAEL	90	1.53%
POLAND	85	1.45%
EGYPT	83	1.41%
NETHERLANDS	80	1.36%
IRAN	70	1.19%
ROMANIA	68	1.16%
RUSSIA	63	1.07%
MEXICO	62	1.06%

3. Prospective Analysis

By adjusting a time series to a life cycle model we are able to predict its future tendency, under the premise that the effects of the external environment won't change (Business as usual).

On those time series that the adjustment to the logistic function cannot be done with an acceptable degree of accuracy (we consider that coefficients of determination lower than 0.75 are unacceptable), we classify their behavior as Saturated o Erratic. If the average between the maximum and the minimum data is ±5% of the series average, we consider Saturation. Otherwise, we classify the series as Erratic.

In the following figures we show the results of such approximations with its coefficient of determination.

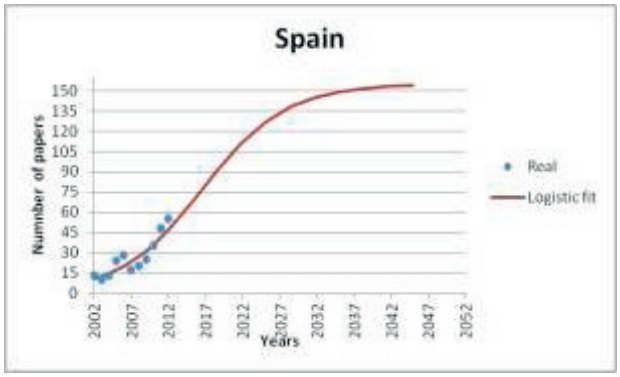

Fig. 1. Evolution of scientific production in Spain. Saturation circa 2040, with approximately 150 papers per year (R2=0.799)

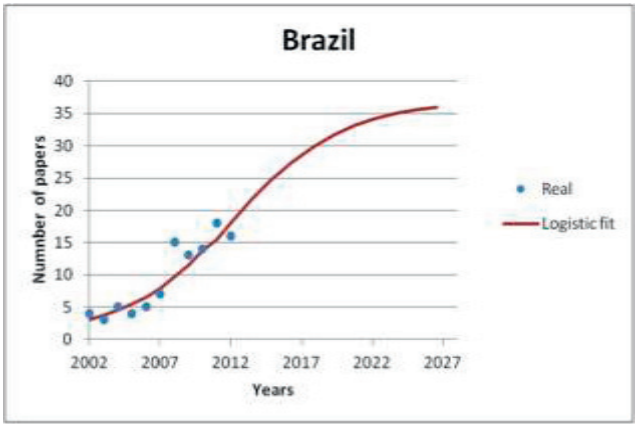

Fig. 2. Evolution of scientific production in Brazil. Saturation circa 2027, with over 35 papers per year (R2=0.869)

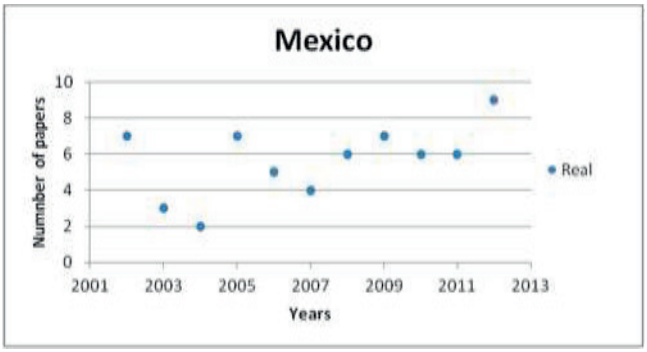

Fig. 3. Evolution of scientific production in Mexico. Erratic behavior, around 5 papers per year.

4. Bibliometric Results

The bibliometric analysis was primarily based on the quantity of published articles, its origin and publishing journal. Where, by origin, we refer to the institution and home address of the author. In this section we present the results of the bibliometric analysis.

The following tables present the top five authors for each of the Ibero-American countries that were analyzed.

Table 2. Top 5 authors with address in Spain (since Berenguel, Cabeza and FernadezIbanez all have the same number of publications, we included them equally),

Authors	No. of papers
MALATO, S	13
BLANCO, J	9
VAZQUEZ, AJ	8
BERENGUEL, M	7
CABEZA, LF	7
FERNADEZIBANEZ, P	7

Table 3. Top 5 authors with address in Brazil,

Authors	No. of papers
MARTINS, FR	9
PEREIRA, EB	9
RUTHER, R	8
TIBA, C	6
ABREU, SL	5

Table 4. Top 5 authors with address in Mexico,

Authors	No. of papers
GONZALESHERNANDEZ, J	5
JARAMILLO, OA	5
ARANCIBIABULNES, CA	4
BANDALA, ER	4
VOROBIEV, YV	4

It is important to note that there were only taken into account the authors who had explicitly established the paper topic as “solar energy”. The majority of the papers fall into the WoS category of “ENERGY & FUELS” as expected.

An interesting component, worth analyzing is the collaborations among countries in these papers. The top four countries that contributed for the total papers for each country are presented.

Table 5. Countries with the highest number of collaborations with Spain.

Countries	No. of papers
GERMANY	20

FRANCE	16
USA	15
ITALY	7

Table 6. Countries with the highest number of collaborations with Brazil.

Countries	No. of papers
USA	8
CHILE	6
GERMANY	5
SPAIN	5

Table 7. Countries with the highest number of collaborations with Mexico.

Countries	No. of papers
USA	7
FRANCE	3
SPAIN	3
ARGENTINA	2
CHILE	2

It is clear that there is an important collaboration with the United States and the Latin-American countries, whereas Spain has a stronger collaboration with the geographically closer Germany and France. France represents a strong collaborator for Mexico, but not for Brazil, whose second strongest collaborator is Chile.

In following tables the top three country pairs are presented. It is important to note that the great majority of the research in this field is done within each country, since most of the papers are found in the first field of each table, the one that accounts for the work done internally.

Table 8. Top collaborations with Spain.

Country collaborations	No. of papers
SPAIN	179
SPAIN:GERMANY	12
SPAIN:FRANCE	11
SPAIN:USA	10

Table 9. Top collaborations with Brazil.

Country collaborations	No. of papers
BRAZIL	77
BRAZI:USA	5
BRAZIL:CHILE	3
BRAZIL:GERMANY	3
BRAZIL:SPAIN	3

Table 10. Top collaborations with Mexico.

Country collaborations	No. of papers
MEXICO	44
MEXICO:SPAIN	3
MEXICO:USA	3
MEXICO:FRANCE	2

Another aspect worth revising is the institutions that produce the papers. The following tables show the most prolific intuitions in each country on “solar energy”.

Table 11. Most prolific Spanish organizations.

Organizations	No. of papers
CIEMAT	44
UNIV ALMERIA	31
CSIC	24

Table 12. Most prolific Brazilian organizations.

Organizations	No. of papers
UNIV SAO PAULO	18
UNIV FED SANTA CATARINA	16
UNIV FED PERNAMBUCO	8

Table 13. Most prolific Mexican organizations.

Organizations	No. of papers
UNAM	25
CINVESTAV	6
IMTA	5

It is noteworthy that in both Spain and Brazil the top three differ of each other in less than ten percent. In Mexico, the most prolific organization UNAM has over three times more than the second most prolific organization Mexican organization, and having six more papers than Brazil’s top one, eventhoug the country’s overall production is greater than Mexico’s. Also, it is interesting that only Brazil’s top three are universities, while in Spain and Mexico, there are also research centers in their lists.

In the following tables we present the authors with the most referenced papers out of the total papers for each country.

Table 14. Most cited authors on Solar Energy in Spain.

Author	No. of references
MALATO, S	47
GULDI DM	26
RODRIGUEZ GP	21

Table 15. Most cited authors on Solar Energy in Brazil.

Author	No. of references
MARTINS, FR	28
PEREZ R	17
RUTHER R	16

Table 16. Most cited authors on Solar Energy in Mexico.

Author	No. of references
JARAMILLO, OA	10
ANIPSITAKIS, GP	7
LUQUE, A	7

It stands out that prolific authors are not necessarily most referenced authors. In both Mexico and Spain, only the most referenced author can be found among the most prolific. And in the case of Brazil, only Perez R is not found among those percentages of contribution on the topic is noteworthy.

5. Citation mining

In this section the results of the analysis done with the text mining of the abstracts of the papers examined are shown. It must be mentioned that the keywords or titles have not been taken into consideration, seeing as how we consider that the relevant information lies within the abstracts.

In table 17 we present the relevant words extracted from the abstracts. There were found over 150 relevant words, we present the top 15 for each country organized by relevance and frequency.

Table 17. Most important (relevant and frequent) words for each of the countries; in **bold** those who appear in all three, in *italics* those who appear in at least two..

Brazil	Mexico	Spain
WATER	**WATER**	**SYSTEM**
SYSTEM	**SYSTEM**	**WATER**
POWER	*FILMS*	*POWER*
RADIATION	**HEAT**	MODEL
HEAT	*TEMPERATURE*	**HEAT**
PHOTOVOLTAIC	FENTON	*RADIATION*
FILMS	HYDROGEN	*THERMAL*
CORROSION	TIO2	COOLING
GENERATION	*THERMAL*	*TEMPERATURE*
COLLECTOR	CELL	DIFFERENT
ELECTRICITY	*MODEL*	AIR

With the same methodology we obtained two word phrases, as presented in table 18.

Table 18. Most important (relevant and frequent) words for each of the countries; in **bold** those who appear in all three, in *italics* those who appear in at least two.

Brazil	Mexico	Spain
WATER HEATING	**PHOTO FENTON**	**PHOTO FENTON**
IN BRAZIL	THIN FILMS	*SOLAR RADIATION*
SOLAR IRRADIATION	*SOLAR RADIATION*	SOLAR THERMAL
RENEWABLE ENERGY	HIGH TEMPERATURE	SOLAR COOLING
SOLAR WATER	HYDROGEN PEROXIDE	PARABOLIC TROUGH
GRID CONNECTED	FENTON REACTION	SOLAR COLLECTORS
PHOTO FENTON	*SOLAR CELL*	PHOTOVOLTAIC SOLAR
ELECTRIC POWER	FRESH WATER	ABSORPTION CHILLER
EXPERIMENTAL RESULTS	PV MODILE	CONCENTRATED SOLAR
HOT WATER	SOLAR CONCENTRATORS	*SOLAR CELL*
WATER HEATING	**PHOTO FENTON**	**PHOTO FENTON**

With this table the outlook of solar energy research gets narrowed down and we can distinguish the lines of research.

Finally, we show the journals where most of the research in these countries is being published in Table 19. The nature of these journals is an indicator of the type of research work done in the countries under study

Table 19. Top 3 journals by country (in the case of Spain we show one more journal, since the last two have the same number of publications).

Spain	Brazil	Mexico
SOL ENERGY	RENEW ENERG	SOL ENERGY
RENEW ENERG	SOL ENERGY	APPL THERM ENG
DESALINATION	ENERG POLICY	SOL ENERG MAT SOL C
INT J HYDROGEN ENERG		

6. Concluding Remarks

The analysis of all the research papers published on journals registered in the Web of Science, with the phrase “solar energy” in its title, abstract or keywords that have at least one author with address in Ibero-America, revealed not only the behavior of the scientific community, but also its strengths and weaknesses. We are able to show the journals where most of their research is being published; the name of the most prolific authors, as well as the most cited; the collaboration with other counties; the strongest areas in the field; and the evolution of the community as a whole. Also, a logistic algorithm to each data series was applied, to make a prospective analysis to different information clusters. Spain has an upward tendency and in the next 30 years it will triple the production on these papers, Brazil has also an upward tendency and in the next 15 years it will double the production of these papers and Mexico has an erratic production. Surprisingly, the photochemical degradation method for organic contaminants, called, photo-Fenton, is a hot topic area in the three countries. This analysis combined with the historical information on the institutional milestones provides a better understanding of the influence of different parameters on the productivity of this particular scientific community.

The previous results allow fostering research areas, collaborations and knowledge transfer strategies between different research groups and leaders, in order to enhance the productivity of this important scientific field.

Acknowledgements

We would like to thank Héctor D. Cortés for developing the initial text mining algorithm and José Abel de la Fuente, for developing the interactive logistic fit algorithm used in this project.

References

[1] Kostoff R., del Río J. A., Humenik J., García E. O. y Ramírez A., Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling, Journal of the american society for information science and technology 52(13):1148-1156. 2001.

[2] del Río, JA, Cortés, HD., La ciencia mexicana en las revistas Nature and Science: la última década. Ciencia, Revista de la Academia Mexicana de Ciencias, 58:61-68. 2007.

[3] Cortés, HD, del Río, JA, García, EO and Robles, M., Web application to profiling scientific institutions through citation mining. Enformatika.14:419-423. 2006.

[4] Ortuno, M, Carpena, P, Bernaola-Galvan, P, Muñoz, E, Somoza, AM., Keyword detection in natural languages and DNA. Europhysics Letters, 57, 759-764. 2002.

[5] Reiss, H., Hammerich, A.D., Montroll, E.W. Thermodynamics of nonphysical systems: formalism and an example (single-lane traffic). J. Stat. Phys. 42: 647-687. 1986.

[6] Russell JM, del Río JA and Cortés HD, Highly Visible Science: A Look at Three Decades of Research from Argentina, Brazil, Mexico, and Spain. Interciencia 2007.

[7] Martinez, M, Seco, RM, Wriedt, K., Futuros de la Universidad: UNAM 2025, UNAM, Mexico, ISBN: 968-842-612-1. 1996.